DOCUMENT RESUME

ED 441 838 TM 030 873

AUTHOR	Garcia, Paul A.
TITLE	Increasing the Rigor of Evaluation Studies of Programs for English Learner Students.
PUB DATE	2000-04-00
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*English (Second Language); *Evaluation Methods; Evaluation Problems; Limited English Speaking; *Program Evaluation; Program Implementation; *Qualitative Research; Secondary Education; *Secondary School Students; Urban Schools

ABSTRACT

This paper outlines important considerations for program evaluators and school administrators when creating evaluation designs, disaggregating student outcome data, and interpreting and reporting achievement data for improved program implementation and informed classroom instruction to ensure treatment of English Learner (EL) students. Evaluation data from a large urban school district serving more than 25,000 EL students are presented to increase understanding and rigor of problems in serving EL students. The discussion focuses on narrowing the limitations of norm-referenced tests through a multiple measures model, overcoming the misguided treatment of EL students as a homogeneous monolithic student population, and the increased value of qualitative research data for program evaluations. The use of a multiple measures model to determine student achievement links student assessments more closely to instruction and curriculum, narrows the limitations of tests, and ensures that EL students are held to the same standards as non-EL students. Surveys of 390 high school students and 882 middle school students demonstrated how EL students felt about their educational experiences. Surveys completed by 102 parents of students in a two-way bilingual immersion program provided qualitative information about parent responses to their children's instruction. (Contains 12 tables, 6 figures, and 13 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

FRESNO UNIFIED SCHOOL DISTRICT
Research, Evaluation and Assessment
Larry Zander, Interim Director

Paul Garcia, Administrative Analyst
pagarcia@fresno.k12.ca.us
457-3839; FAX 457-3794

ED 441 838

Increasing the Rigor of Evaluation Studies of Programs for English Learner Students

Paul A. Garcia, Ed. D.
Fresno Unified School District

TM030873

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
P. Garcia
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 2000

BEST COPY AVAILABLE

Increasing the Rigor of Evaluation Studies of Programs for English Learner Students

Introduction

Standards based educational reform efforts that include high stakes testing for school accountability, to eliminate social promotion, and as criterion for student retention are cause for serious concern in districts with large numbers of English Learner (EL) students. In California, the *Stanford, Ninth Edition (SAT9)* standardized norm referenced test is required in school accountability models and to determine student performance. Since student assessments are typically used for program evaluations, when applied to evaluations of programs serving EL students, the problematic nature of standardized norm referenced tests warrants careful attention. Standardized norm referenced tests fail to accurately assess the academic abilities of EL students (August and Hakuta, 1997), traditional testing accommodations are often inappropriate for EL students (Rivera et al., 2000), and there is a paucity of non-English language assessments (Garcia et al. 2000). Therefore, careful examination is necessary of appropriate assessments and proper reporting procedures used to determine EL student achievement and second language acquisition.

The pervasive use of high stakes testing to increase student achievement creates an interesting paradox. How can the increased use of standardized tests, typically related to unwarranted assaults on the educational experiences of language minority students, help reverse historical inequities in schooling? If major increases in student achievement are to occur, traditional viewpoints and practices concerning the collection, interpretation, and reporting of achievement data for EL students must improve. In other words, schools are challenged to use achievement data not to sort and select students, but to create school cultures that guide school practices and inform instruction. The answer to the paradoxical question of how schools can

effectively use achievement data to improve the educational experiences of language minority students lies in the critical examination of data not to change students with more schooling, but to transform the school environment.

Increased rigor of program evaluations can mitigate harmful effects of high stakes testing on language minority students and prevent the exclusion of large numbers of students from participation in school reform. The purpose of this paper is to outline important considerations for program evaluators and school administrators when creating evaluation designs, disaggregating student outcome data, and interpreting and reporting achievement data for improved program implementation and informed classroom instruction. Valid, accurate, and inclusive assessments can improve access to content standards (equity) and increase academic achievement (excellence) among EL students. Evaluation data from a large urban school district serving over 25,000 EL students will be presented to increase understanding and rigor of evaluations of programs serving EL students. Discussion will focus on narrowing the limitations of norm referenced tests through a multiple measures model, overcoming the misguided treatment of EL students as a homogeneous monolithic student population, and the increased value of qualitative research data for program evaluations.

Narrowing the Limitations of Norm Referenced Tests in a Multiple Measures Model

Limitations of standardized norm referenced tests

An important goal of programs serving EL students is to close the achievement gap between EL and non-EL students. EL students represented a large percentage of tested students in the four largest school districts in California; Los Angeles (36%), San Diego (24%), Santa Ana (63%), and Fresno (29%). California statewide test results indicate EL students scored about 30 percentile ranks below non-EL students in reading and language, and 25 percentile

ranks lower in mathematics. Accurate determination of the achievement levels of EL students is often confounded by the language barrier of English language norm referenced tests. Including EL student test scores in student academic profiles can have unintended and harmful consequences. The stakes are high when schools use test scores as criteria for grade level retention, summer school, assignment to remedial or accelerated programs (GATE, Advanced Placement, and honors), high school graduation, or as indicators of school performance.

Among many EL students, test results often reflect English language proficiency more than academic achievement (August & Hakuta, 1997). In some cases, test results are inappropriately used as indicators of English language proficiency and consequently, determinants of program placement (Gonzalez et al., 1996). The misguided use of standardized norm referenced tests for EL students is in part attributed to confusion over language arts as a content area (inference, comprehension, literary concepts, and reading abilities) and English language proficiency as a second language (beginning to advanced levels). The misunderstanding is over the difference between literacy skills and language proficiency, or content versus CALP (Cognitive Academic Language Proficiency; Cummins, 1989).

Objective measures such as standardized norm referenced tests (NRT) are often too simplistic to assess cause and effect relationships between instructional practices and student learning (Gamoran & Berends, 1987). Teacher and student interactions are too complex to be captured by standardized tests. Moreover, content objectives of NRTs are not designed to match the curriculum students experience in the classroom. Emphasis on requiring students to meet curricular and performance standards has led many to believe NRTs can accurately assess proficiency levels in content areas.

As indicated in **Table 1**, there is a positive relationship between SAT9 test results and

English language proficiency. Students with higher levels of English language proficiency (ELD IV-V) had higher mean Normal Curve Equivalent (NCE) scores: Spanish language students exhibited higher test scores on the Spanish language norm referenced test, (SABE/2), compared to SAT 9. The scores varied by as much as 24 NCEs between the two tests.

Table 1
Comparison of *SAT9* and *SABE/2* Mean NCEs for EL Students
Completing Both Tests, By ELD Level

	ELD I-III		ELD IV-V	
READING	*Mean*	*Frequency*	*Mean*	*Frequency*
SABE2	46.3	1925	48.6	224
SAT9	22.1	1925	30.6	224
LANGUAGE				
SABE2	46.1	1996	50.3	222
SAT9	25.4	1996	33.3	222
MATHEMATICS				
SABE2	44.6	2077	52.0	235
SAT9	30.0	2077	39.2	235

Multiple Measures

The use of a multiple measures model to determine student achievement links student assessments more closely to curriculum and instruction, narrows the limitations of NRTs, and ensures EL students are held to the same high standards as non-EL students. Results of standardized tests for EL students are highly suspect and difficult to interpret when compared to normed populations not representative of ethnically and linguistically diverse student populations. The use of multiple measures increases reliability of assessments to determine EL student achievement levels for informed instruction and program evaluations. In FUSD, multiple measures include criterion reference tests, literacy and mathematics student profiles, and the Spanish language NRT, SABE/2.

District criterion referenced test

Currently, SAT9 is not aligned with California English language arts or mathematics content standards. Therefore, this year FUSD developed and implemented a criterion referenced test to assess students in grades 2 through 9 on district standards in language arts and mathematics. The Assessment of Baseline Curriculum (ABC) is an English language test and is administered twice a year to determine student gains toward proficiency on the standards. Students at lower levels of English language development (ELD) I, II, and III were exempted from the ABC testing program, while students at higher English language proficiency levels (ELD IV and V) were included. EL students at ELD I-III participated in an alternative assessment on district standards. Later, when ABC results were used as one criterion for student retention, the decision to exempt some EL students prevented the inappropriate retention of students.

As indicated in **Table 2**, English only students were more likely to achieve proficiency levels on the ABC test.

Table 2
Percent of English Learner and English Only Students
Achieving Proficiency on Reading, Language, and Mathematics Standards

	English Learners	English Only
Reading		
Grades 3-6	30%	56%
Grades 7-9	16%	56%
Language		
Grades 3-6	37%	62%
Grades 7-9	24%	65%
Mathematics		
Grades 3-6	44%	56%
Grades 7-9	24%	52%

Literacy and mathematics student profiles

The decision to use an alternative assessment to assess EL students at ELD levels I-III on district standards was based on equity, instructional, and research based issues. Teachers completed a computer scanner form for each student indicating proficiency levels on the seven language arts and six mathematics standards in grades kindergarten through six. Teacher observations and student work were used to complete the scanners. The alternative assessment provided important achievement data for monitoring EL student progress on content standards and reinforced the district's goal to raise achievement levels for all students. Spring assessment results are expected to indicate academic growth and English language development gains. Many EL students are expected to make sufficient ELD gains during the academic year to either transition into the ABC testing program or meet increased proficiency levels on the content standards.

Primary language norm referenced test

To determine achievement gains among Spanish language students, test scores are presented for students who took the *SABE/2* three consecutive years. For the purpose of longitudinal analysis, student scores were aggregated to determine Spanish language academic skills, irrespective of instructional program placement or redesignation status (August and Hakuta, 1997). Examination of results indicate *SABE/2* test scores improved in most grade levels between the spring of 1997 and spring 1999. Of ten grade levels, six experienced gains in reading (**Figure 1**), five in language (**Figure 2**), and six in mathematics (**Figure 3**). Secondary students in grades 7 through 11 experienced substantial gains in reading, language, and mathematics.

SABE/2 Mean NCE Scores for 1997, 1998, and 1999 by Grade
Reading

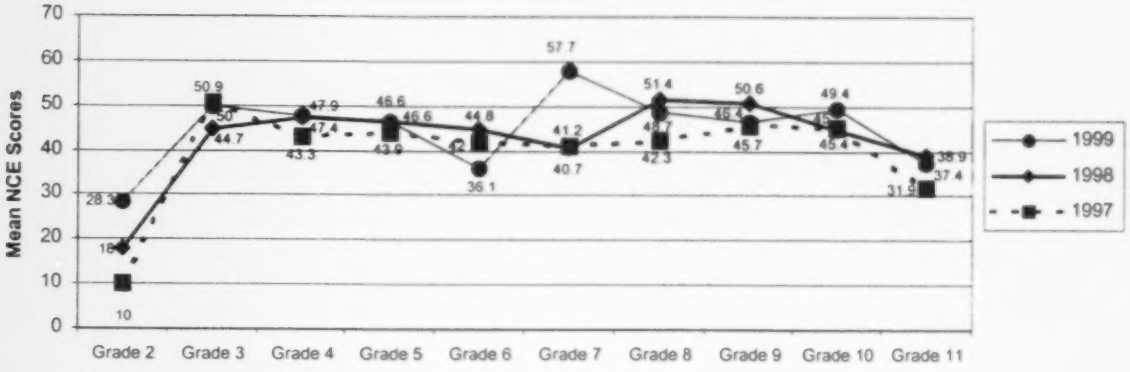

Fig. 1

SABE/2 Mean NCE Scores for 1997, 1998, and 1999 by Grade
Language

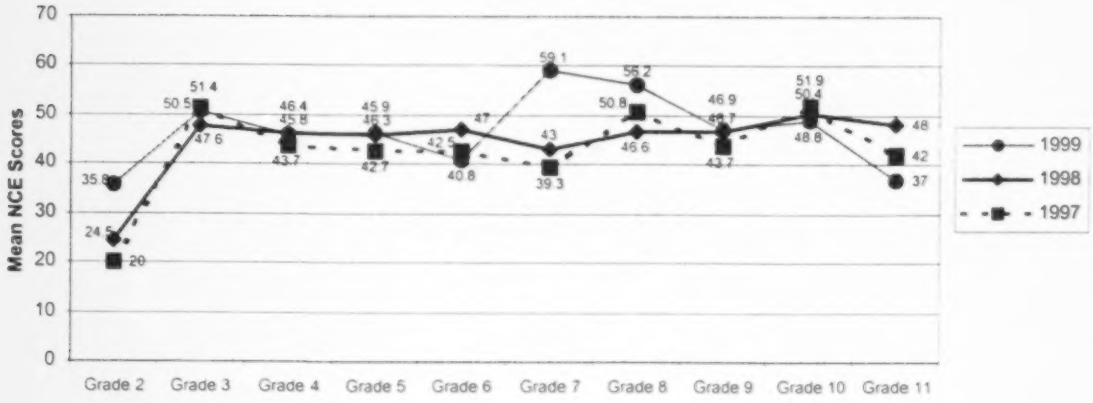

Fig. 2

SABE/2 Mean NCE Scores for 1997, 1998, and 1999 by Grade
Mathematics

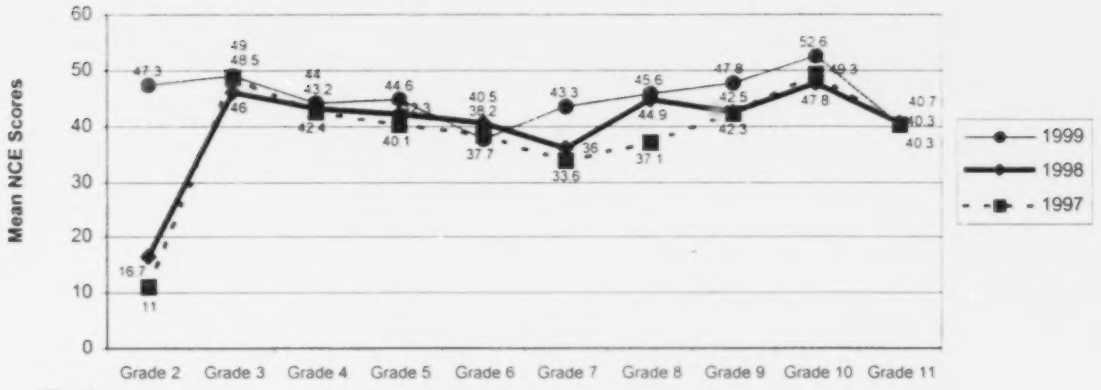

Fig 3

Like English-only students, EL students are expected to meet grade level standards on a norm-referenced test if one exists in their primary language. Therefore, *SABE/2* was used for Spanish language students. As indicated in **Table 3**, a higher percentage of 7th and 8th grade students met grade level standards in reading (48%) and language (58%) by scoring at or above the 50th percentile. A large percentage of students in grades 2 through 6 met grade level standards in mathematics (44%). A higher percentage of students met grade levels standards in 1999 than 1998 across all three test areas in grades 2 through 6, and 9 through 11.

Table 3
Percent of EL Students Scoring At or Above the 50th Percentile on *SABE/2*, 1998 and 1999

	Reading		Language		Mathematics	
	1998	1999	1998	1999	1998	1999
Grades 2-6	36% (618)	41% (899)	38% (638)	43% (949)	32% (551)	44% (963)
Grades 7-8	49% (113)	48% (103)	41% (94)	58% (126)	34% (79)	32% (68)
Grades 9-11	36% (101)	44% (133)	40% (109)	43% (111)	32% (88)	37% (102)

One of the goals in the *FUSD Master Plan for Limited English Proficient Students* (FUSD, 1996) is to increase English and primary language proficiency. Students experience an instructional program that promotes additive bilingualism by gaining proficiency in English without losing primary language skills. SABE/2 results provide evidence that Spanish language students are learning English, but not the expense of losing Spanish language academic skills.

English Learners Are Not a Homogeneous Student Population

EL students are often treated as a homogeneous monolithic population with common historical and educational experiences. For example, California requires the posting of SAT9 scores on the Internet, but groups all EL students together. Yet others have found language minority groups to experience diverse educational experiences (Ogbu, 1988) and expectations (Matute-Bianchi, 1986). EL student educational experiences can vary within schools and school districts due to the differential availability of primary language instructional materials, teachers, or paraprofessionals; or differential placement in instructional programs. Interpretation of test results is improved when student achievement data is disaggregated by language group membership, English language proficiency, and participation in bilingual programs. Disaggregation of achievement data can help determine effects of instructional program on diverse student populations and improve identification of program features related to increased student achievement.

Language group differences

As indicated in **Figures 4-6**, SAT 9 test scores differed substantially among language groups. Examination of *SAT9* test scores among the five largest language groups in FUSD, indicate Vietnamese students exhibited the highest test scores in reading (Fig. 4), language (Fig.

5), and mathematics (Fig. 6). However, caution should be exercised when interpreting the data due to the small number of Vietnamese students tested. All language groups had higher mean NCE scores in mathematics. Among the two largest language groups, Spanish language students had higher scores in reading and Hmong language students had higher scores in language and mathematics. Wide differences in test scores among language groups warrants close attention to reasons some EL students more than others have high achievement levels.

SAT9 Mean NCE Scores by Language and Grade, ELD IV-V
Reading, 1998 and 1999

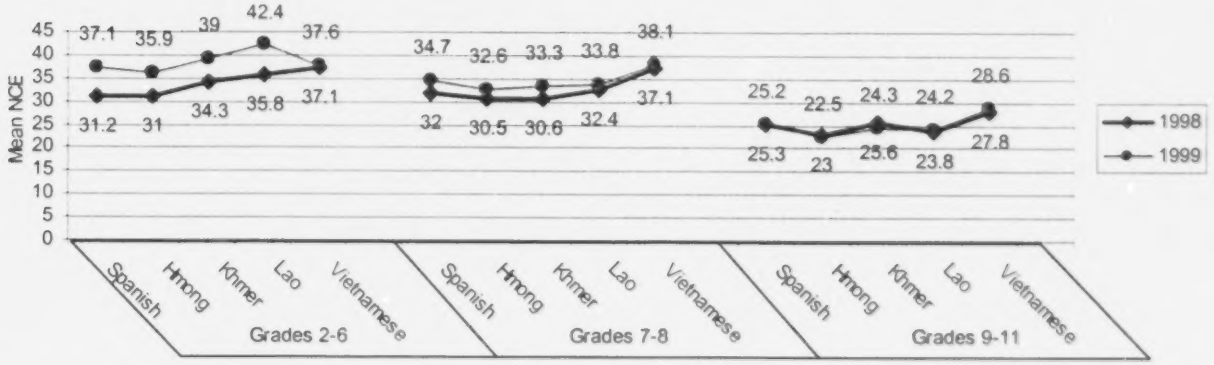

Fig. 4

SAT9 Mean NCE Scores by Language and Grade, ELD IV-V
Language, 1998 and 1999

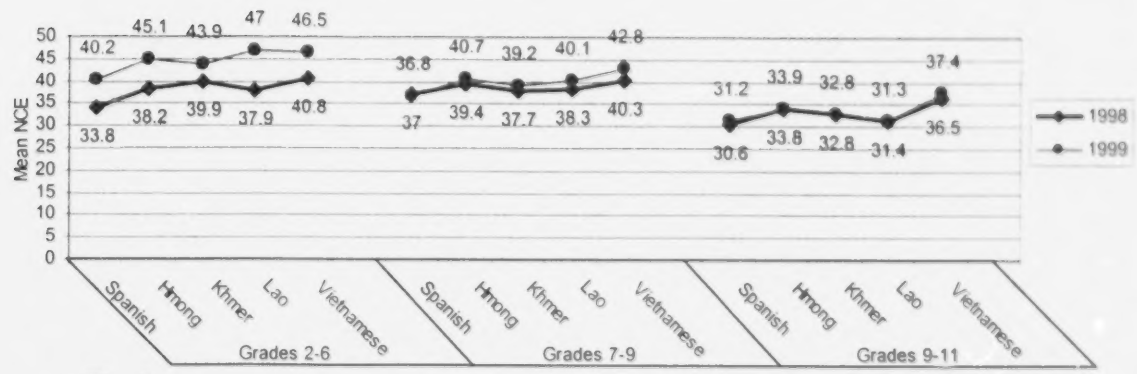

Fig. 5

BEST COPY AVAILABLE

SAT9 Mean NCE Scores by Language and Grade, ELD IV-V
Mathematics, 1998 and 1999

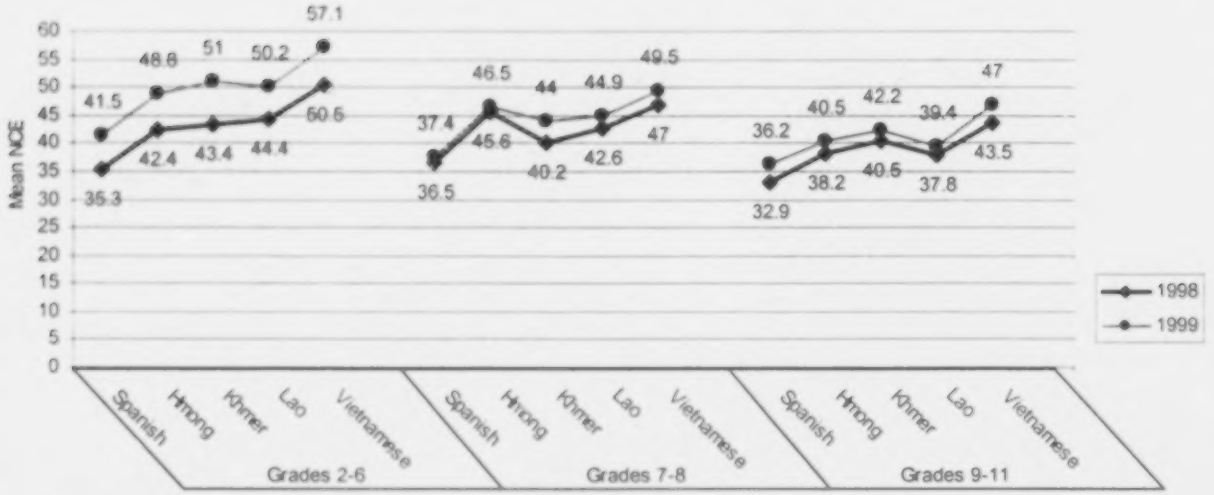

Fig. 6

English language proficiency differences

This section presents evidence of the effect English language proficiency had on SAT 9 test scores. Students at higher levels of English language development (ELD) levels IV and V, performed better than students at ELD I, II, and III. Test data in **Tables 4-6** indicate wide disparities in mean NCE scores between ELD I-III and ELD IV-V students in reading (6 NCEs), language (8 NCEs), and mathematics (8 NCEs). In language, students at ELD levels I-III and ELD levels IV-V made at least 4 mean NCE gains in all grade categories (**Table 5**). ELD I-III students in grades 7 through 8, and 9 through 11 made 5 mean NCE gains. Gains of at least 5 NCEs are necessary to close the achievement gap between EL and English-only students (Thomas and Collier, 1996). Redesignated students and English-only students made about 4 mean NCE gains in grades 7 through 8, and 9 through 11. All student groups made at least 2

mean NCE gains in mathematics across all grade categories (**Table 6**).

Table 4

SAT9 Mean NCE Scores By ELD Level, English-Only, and Redesignated Students (FEP-R)
Reading, Grades 2-11

	ELD I-III		ELD IV-V		FEP-R		English-Only	
	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean
Grades 2-6	22.0 (6768)	24.9 (6960)	31.6 (2921)	34.3 (3129)	49.6 (855)	51.7 (762)	41.7 (15310)	43.8 (16789)
Grades 7-8	15.0 (918)	17.9 (1001)	28.1 (2036)	30.6 (1980)	47.6 (900)	50.9 (832)	45.0 (5639)	46.1 (6059)
Grades 9-11	10.4 (567)	11.0 (513)	20.9 (2489)	22.5 (2835)	39.8 (1333)	41.7 (1211)	42.0 (6852)	42.5 (7377)
Total	20.4 (8253)	23.2 (8474)	27.1 (7446)	29.2 (7944)	44.8 (3,088)	42.2 (2805)	42.4 (27801)	43.9 (30225)

Light shaded area denotes gains of at least 2 NCEs. Dark shaded area denotes gains of least 4 NCEs.

Table 5

SAT9 Mean NCE Scores By ELD level, English-Only, and Redesignated Students (FEP-R)
Language, Grades 2-11

	ELD I-III		ELD IV-V		FEP-R		English-Only	
	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean
Grades 2-6	23.8 (7506)	27.9 (7202)	35.0 (3080)	39.0 (3147)	55.4 (866)	58.4 (755)	40.3 (16229)	44.2 (16993)
Grades 7-8	19.0 (983)	24.2 (1012)	31.5 (2103)	35.9 (1967)	51.3 (905)	55.9 (812)	43.3 (5858)	47.4 (5965)
Grades 9-11	15.4 (719)	21.3 (517)	27.5 (2673)	31.6 (2805)	46.0 (1362)	50.4 (1196)	41.9 (7304)	46.8 (7270)
Total	22.6 (9208)	27.0 (8731)	31.5 (7856)	35.6 (7919)	50.1 (3,133)	54.2 (2763)	41.3 (29391)	45.5 (30228)

Light shaded area denotes gains of at least 2 NCEs. Dark shaded area denotes gains of least 4 NCEs.

Table 6

SAT9 Mean NCE Scores By ELD Level, English-Only, and Redesignated Students (FEP-R)
Mathematics, Grades 2-11

	ELD I-III		ELD IV-V		FEP-R		English-Only	
	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean	1998 Mean	1999 Mean
Grades 2-6	27.7 (7506)	32.5 7336	39.1 (3080)	42.3 (3220)	57.1 (866)	61.3 (769)	41.5 (16229)	45.1 (17245)
Grades 7-8	26.6 (983)	30.4 (1029)	34.7 (2103)	38.6 (1980)	51.7 (905)	55.0 (834)	42.5 (5858)	46.3 (6124)
Grades 9-11	25.6 (719)	29.2 (570)	34.3 (2673)	38.1 (2876)	51.1 (1362)	54.7 (1205)	45.0 (7304)	49.4 (7416)
Total	27.5 (9208)	32.0 (8936)	36.3 (7856)	39.9 (8076)	53.0 (3,133)	56.6 (2808)	42.6 (29391)	46.4 (30785)

Light shaded area denotes gains of at least 2 NCEs. Dark shaded area denotes gains of least 4 NCEs.

SAT9 test scores of ELD IV-V students are a more accurate determination of EL student achievement since English language proficiency is less of a factor. While ELD IV-V students trailed English-only students in reading, language, and mathematics in all grade categories, the gap varied by grade and test area (See **Table 7**). The largest gap occurred in reading among grades 9 through 11 (20 mean NCEs), while the achievement gap was only 3 mean NCEs in mathematics among students in grades 2 through 6.

Table 7

Summary of SAT9 Mean NCE Scores for English Learner (EL)*
and English-Only (E-O) Students, 1998 and 1999

	Mean NCEs								
	Grades 2-6			Grades 7-8			Grades 9-11		
	EL	E-O	Difference	EL	E-O	Difference	EL	E-O	Difference
Reading									
SAT 9 1998	31.6	41.7	10.1	28.1	45.0	16.9	20.9	42.0	21.1
SAT9 1999	34.3	43.8	9.5	30.6	46.1	15.5	22.5	42.5	20.0
Language									
SAT 91998	35.0	40.3	5.3	31.5	43.3	11.8	27.5	41.9	14.4
SAT9 1999	39.0	44.2	5.2	35.9	47.4	11.5	31.6	46.8	15.2
Mathematics									
SAT 9 1998	39.1	41.5	2.4	34.7	42.5	7.8	34.3	45.0	10.7
SAT9 1999	42.3	45.1	2.8	38.6	46.3	7.7	38.1	49.4	11.3

* ELD level IV and V students only.

Instructional program comparisons.

SAT9 Test Results

As suggested above, English language proficiency is strongly related to *SAT9* test scores. In part, the relationship is attributable to the students' instructional program. Students in *Structured English Immersion (SEI)* are at the early stages of English language acquisition and receive instruction that is overwhelmingly in English. Students in *primary language* classes (*L1*) receive Spanish language instruction, and students in *mainstream* classes have reasonable fluency in English and receive a form of sheltered English instruction in content areas.

Data presented in **Table 8** suggest students in grades 2 through 6 in *SEI* or *primary language* classrooms had similar mean NCE scores and gains in reading, language, and mathematics. Among students in grades 9 through 11, students in *primary language* classes had somewhat higher mean NCE scores than students in *SEI* classes in reading (*SEI*, 10.5; *L1*, 12.5) and language (*SEI*, 20.6; *L1*, 22). The very small number of students with *SAT9* scores in *primary language* classes in grades 7 through 8, precludes valid comparisons. *Mainstream* students had the highest mean NCE scores in reading, language, and mathematics. In summary, evidence supports other research (Ramirez, Yuen, and Ramey, 1991) that primary language instruction does not lead to decreased achievement levels among EL students.

Table 8

SAT9 Mean NCE Scores for English Learner Students in Bilingual, Structured English Immersion (SEI), and Mainstream Instructional Programs, 1998 and 1999

	Grades 2-6			Grades 7-8			Grades 9-11		
	SEI	Bilingual	Mainstream	SEI	Bilingual	Mainstream	SEI	Bilingual	Mainstream
Reading	*(3284)*	*(520)*	*(2348)*	*(821)*	*(2)*	*(1770)*	*(100)*	*(216)*	*(2370)*
1999	22.5	23.3	33.4	18.4	16.1	30.8	10.5	12.5	23.2
1998	20.5	22.2	29.1	17.6	17.2	28.4	11.3	12.7	23.9
Language	*(3475)*	*(574)*	*(2437)*	*(837)*	*(2)*	*(1752)*	*(97)*	*(213)*	*(2340)*
1999	27.5	27.3	38.4	24.7	12.9	35.9	20.6	22.0	32.1
1998	23.9	24.2	32.6	23.0	16.1	34.5	18.7	19.3	31.2
Mathematics	*(3676)*	*(672)*	*(2541)*	*(873)*	*(2)*	*(1826)*	*(121)*	*(263)*	*(2494)*
1999	30.8	30.3	41.8	30.8	17.9	38.8	30.3	30.1	38.7
1998	26.9	27.4	36.6	28.2	16.3	37.7	26.0	27.0	35.3

Light shaded area denotes gains of at least 2 NCEs. Dark shaded area denotes gains of least 4 NCEs.

SABE/2 Test Results

Determination of program effectiveness must include such factors as length of student participation in bilingual education programs, and consistency and coherency of program implementation. The instructional program provided to EL students should not be omitted from data analysis. In 1999, 11 percent of FUSD students received *primary language* instruction, a decrease from 17 percent in 1998. About 5 percent of EL students were in classrooms where appropriately credentialed teachers were unavailable. As indicated in **Table 9,** SABE/2 results suggest students enrolled in *primary language* classes out performed students in *Structured English Immersion* classes. The relatively small number of students in *mainstream* classes with higher levels of English language proficiency, also scored higher than *Structured English Immersion* students. Test results suggest close attention to students in *Structured English Immersion* classes may be warranted to prevent deficits in academic skills assessed in Spanish.

Table 9

SABE/2 Mean NCE Scores for English Learner Students in Bilingual, Structured English Immersion (SEI), and Mainstream Instructional Programs, 1999

	Bilingual	SEI	Mainstream
Reading			
Grades 2-6	47.8 (1319)	39.6 (547)	47.6 (85)
Grades 7-8	38.4 (24)	52.6 (179)	50.6 (14)
Grades 9-12	45.7 (239)	39.1 (26)	40.2 (18)
Language			
Grades 2-6	47.9 (1308)	38.8 (544)	51.4 (86)
Grades 7-8	51.0 (24)	57.2 (178)	62.3 (14)
Grades 9-12	45.8 (211)	31.3 (18)	40.9 (14)
Mathematics			
Grades 2-6	47.0 (1320)	39.5 (549)	53.9 (85)
Grades 7-8	34.5 (24)	42.0 (179)	53.2 (13)
Grades 9-12	44.7 (216)	31.9 (23)	40.9 (16)

Qualitative Data

Evaluation designs of bilingual programs typically rely heavily on quantitative data (standardized test scores, redesignation rates, and gains in language development) to document student progress. Little attention is given to the curricular experiences of EL students including access to a rigorous and challenging curriculum, student perceptions toward second language acquisition, or student career and educational aspirations. The use of such qualitative data to determine program effects makes evaluation designs more rigorous in nature and comprehensive in scope. Qualitative research methods can aid the interpretation of quantitative data by bridging the explanatory gap between student outcome data and student learning experiences.

Classroom observations, teacher interviews, parent surveys, and the inclusion of student voice (Olsen & Jaramillo, 1999) increases contextual understanding of the relationship between achievement gains and instructional practices. Gaining English language proficiency is an important goal of all programs serving EL students. The experience of second language acquisition is important when interpreting student outcome data. For example, instruction in

English language development classes may not meet the needs of students. One student described the remedial and unchallenging instruction found in ELD classes.

> ESL (English as a Second Language) students get the lowest classes possible, like wood shop. When you get into ESL, you end up in ESL. I think that's discrimination. They treat you like little kids.

Student Survey Results

To determine how students feel about their schooling, a sample of secondary school EL and English-only students were surveyed to assess school experiences, educational goals, and aspirations. Survey results are presented for two high schools (390 students) and three middle schools (881 students). Classrooms were selected to represent diverse language groups and varying levels of English language proficiency. Positive school experiences are indicated by a relatively high percentage of students responding either *strongly agree or agree* to question items. Question items were designed to assess three areas: educational plans, student guidance, and school participation.

Educational plans. Survey results indicate responses toward educational plans did not differ between students in grades 7 through 8, and 9 through 12. As indicated in **Table 10,** a similarly high percentage of students *planned to complete high school (*ELD I-III, 97%; ELD IV-V, 99%; English-only 99%) and *attend and graduate from a college or university* (ELD I-III, 89%; ELD IV-V, 92%; English-Only, 93%). A large percentage of students in ELD I-III indicated an interest in *attending a trade or specialized training program after high school* (grades 7-8, 70%; grades 9-12, 65%). English only students were less likely to indicate receiving *help from teachers or counselors in planning their education after high school* (grades 7-8, 52%; grades 9-12, 38%).

Table 10
Secondary Student Survey: Educational Plans

	Strongly Agree/Agree Grades 7-8			Strongly Agree/Agree Grades 9-12		
	English-Only (311)	ELD I-III (269)	ELD IV-V (301)	English-only (40)	ELD I-III (242)	ELD IV-V (108)
I plan to complete high school.	99%	97%	99%	98%	99%	98%
I plan to attend and graduate from a college or university.	93%	89%	92%	85%	88%	90%
I plan to attend a trade/specialized training program after high school.	48%	70%	58%	46%	65%	48%
Teachers and/or counselors are helping me plan my education after high school graduation.	52%	74%	64%	38%	69%	56%
My school classes will help in my adult life.	90%	89%	90%	85%	91%	90%

Academic guidance. As indicated in **Table 11**, a higher percentage of students at ELD I-III, than their English only counterparts, indicated *teachers and counselors were helpful* (grades 7-8, 74%; grades, 9-12, 69%). ELD I-III students were also more likely to indicate *they were doing as well in school as they would like to* (grades 7-8, 83%; grades 9-12, 90%).

Table 11
Secondary Student Survey: Academic Guidance

	Strongly Agree/Agree Grade 7-8			Strongly Agree/Agree Grade 9-12		
	English-Only (311)	ELD I-III (269)	ELD IV-V (301)	English-only (40)	ELD I-III (242)	ELD IV-V (108)
When I have a problem I can find help at school.	74%	78%	76%	53%	83%	70%
Teachers and counselors are helpful.	52%	74%	64%	38%	69%	56%
I find it easy to follow the school rules.	76%	75%	80%	75%	90%	87%
I ask my teachers/counselors for help when I need it.	72%	85%	81%	73%	88%	79%
I am doing as well in school as I would like to.	55%	83%	74%	45%	90%	70%

School participation. A larger percentage of ELD I-III students in grades 7 through 8 (74%) and 9 through 12 (79%) indicated *their parents enjoyed visiting their school* (**Table 12**). A small percentage of English only students *encouraged their parents to participate in school-*

BEST COPY AVAILABLE

parent activities in grades 7 through 8 (43%) and 9 through 12 (28%). A relatively smaller percentage of all students indicated *participation in extra-curricular activities*.

Table 12
Secondary Student Survey: School Participation

	Strongly Agree/Agree Grade 7-8			**Strongly Agree/Agree** Grade 9-12		
	English-Only (311)	ELD I-III (269)	ELD IV-V (301)	English-only (40)	ELD I-III (242)	ELD IV-V (108)
My parents enjoy visiting my school.	65%	74%	70%	38%	79%	62%
I encourage my parents to participate in the school's parent activities.	43%	62%	59%	28%	72%	44%
Most students in this school are friendly.	68%	49%	64%	60%	70%	71%
I have friends at school.	98%	96%	99%	98%	95%	96%
I participate in one or more—sports, clubs, music, yearbook, cheer/pep and/or drama.	45%	44%	48%	48%	43%	37%

Results from the secondary student survey suggest it is important to consider the English language proficiency level of students to increase understanding of school experiences and educational goals among EL students.

Parent Survey Results

Parents of students participating in a two-way bilingual immersion program were surveyed to assess their opinions of the program. Of the 119 parents surveyed, 102 responded for a response rate of 86 percent. Survey results indicated many parents were very supportive of their children learning two languages. Parent responses are divided into two categories:

1) Parents indicated bilingual skills would help their children have improved educational and employment opportunities, and 2) Learning a second language would increase student self-esteem.

Improved educational and employment opportunities:

It's very positive. This language will help her in the workforce when she grows up.

Me siento muy contenta que mis hijos están aprendiendo un segundo idioma ya que es muy importante para ellos en el futuro.
[I am very content that my children are learning a second language and it is very important for their future.]

I am happy that my daughter has the opportunity to learn a second language. I am confident that doing so will amount to a personal asset. That will benefit her and others throughout her life.

I think a second language is important for a good job in the future.

Pues me siento muy satisfecha de que mi hija este aprendiendo un segundo idioma, para que en un futuro valga por dos.
[Well, I feel very satisfied that my daughter is learning a second language so that in the future she is valued (for two).]

I feel it's very important especially now in these days. It's so very important for job skills. There's a big demand for people with two languages. It really opens the door for jobs for our children.

Learning a second language would increase student self-esteem:

Me siento contenta porque pueden desenvolverse por se mismos en los dos idiomas porque en mi casa siempre se habla el Español. Y quisiera que nunca dejaran sus raises aunque vivan en un país diferente.
[I feel content because they can be confident in two languages, in my home we only speak Spanish. I want them to never forget their background, although they may live in another country.]

My son has great-grandparents and grandparents who speak Spanish fluently. I don't speak Spanish fluently and this is a great opportunity for him to learn. This can only help him as his education continues into high school and college. I am pleased since she is Hispanic and should speak Spanish, although I, myself, (sic) do not.

Me siento feliz que me mi hija sabe leer y escribir. Cuando va algún lado y me pregunta unas cosas y ella me dice lo que me esta preguntado y ella contesta. Estoy feliz porque ya sabe.
[I feel happy my daughter knows how to read and write in Spanish. When we go places and I am asked questions, my daughter translates for me. I am happy because she understands.]

Bueno, primero que nada, felicidades por tratar de conservar la herencia hispana. En segundo lugar quiero mostrar la inquietud que siento al ver que mis niñas no dominan el Inglés como lo hacen otros niños de su edad, que son hijos de padres Latinos, pero que tienen hermanos que dominan el idioma Inglés.
[First of all, congratulations for trying to conserve the Hispanic heritage. Second, I would like to express my anxiety about my girls not having command of the English language like other kids their age who have Latino parents, but have siblings that are dominant in English]

Bueno lo estoy contenta que hable dos idiomas Español y Inglés y que aprenda a escribirlos también y qūe por aprender Inglés y no olviden el Español. Que bueno que les enseñan los dos idiomas atentamente.
[I am glad that my child speaks both Spanish and English and is able to write both languages. It is good that both languages are being taught and that the Spanish language should not be forgotten.]

Conclusion

State and local efforts to implement standards based educational reform have centered on developing and administering a wide range of student assessments (Linn, 2000). Student assessments are typically incorporated into program evaluations, therefore, careful consideration is warranted when determining the appropriate use of assessments and interpretation of results for evaluation of programs serving EL students. This report has described one district's efforts to implement fair, accurate, and inclusive student assessments for large numbers of EL students. Careful attention was made to ungroup achievement data for the improvement of reporting procedures and interpretation of results.

A number of issues concerning the assessment of EL students for program evaluations surfaced in this report:

1) The process of determining appropriate assessments for EL students was as informative as the results.
2) Careful attention to the appropriate use of assessments can prevent EL students from experiencing unnecessary retention or instructional interventions.
3) English language assessments understate EL students' academic skills that may otherwise be expressed in primary language assessments.
4) An evaluation design that includes qualitative data and multiple measures increases the rigor of program evaluations of services for EL students.

Recommendations

- Evaluation designs of bilingual education programs should incorporate qualitative and quantitative data to better assess effects of instructional practices on student learning and to identify effective program features.

- The treatment of language minority students as a homogeneous monolithic student population prevents illumination of student outcome data that could shed light on student academic or language development needs appropriate to each language group.

- Student voice is a valuable form of evaluation data that can help explain student outcomes. It can assist interpretation of the complex relationship between learning subject content and the experience of second language acquisition.

- The disaggregation of achievement data by English language proficiency increases meaningful interpretation and prevents harmful consequences for large numbers of EL students.

- Careful consideration should be given when testing students in a language that differs from the language of instruction. It is unfair to expect EL students to compete with English language counterparts without exposing them to the same curriculum.

- Increased rigor of program evaluations will have research and public policy implications for closing the achievement gap between EL and English only students.

- Establish a district policy concerning the testing of EL students.

- Provide the opportunity for EL student testing accommodations when appropriate.

- Reports of test results to the school community should ungroup students at lower levels of English language proficiency.

- Inform all parents of the right to exempt students from testing when applicable (i.e. SAT 9). Non-English language standardized norm referenced and criterion referenced tests should be administered when appropriate

- Careful consideration of the conclusions and implications presented in this report can mitigate effects that contribute to the educational neglect of language minority students.

References

August, D. & Hakuta, K. (Eds.). (1997). Improving schooling for language minority children: A research agenda. Washington, DC: National Academy Press.

Cummins, J. (1989). Empowering minority students: A framework for intervention.California Association for Bilingual Education.

Fresno Unified School District Master Plan for Limited English Proficient Students, 1995-2000. (1996). Fresno Unified School District: Fresno, CA.

Gamoran, A. & Berends, M. (1987). *The effects of stratification in secondary schools: Synthesis of survey and ethnograghic research.* Review of Educational Research, 57, 415-435.

Garcia, E. E.; Casimir, M.; Irminger, X.S.; Wiese, A.M.; and Garcia E.H. (2000). Authentic literacy assessment development: An instructional based assessment that is responsive to linguistic and cultural diversity. Paper presented at National Association for Bilingual Education, 2000: San Antonio.

Gonzalez, V.; Brusca-Vega, R.; and Yawkey, T. (1996). Assessment and instruction of culturally and linguistically diverse students with or at-risk of learning problems: From research to practice. Allyn & Bacon: Needham Heights, MA.

Kohn, A. (1999). The schools our children deserve. Boston: Houghton Mifflin

Linn, R. L. (2000). Assessments and accountability. Educational Researcher 29 (2), 4-16.

Matute-Bianchi, M. (1986). *Ethnic identities and patterns of school success among Mexican descent and Japanese students in a California high school: An ethnographic analysis.* American Journal of Education.,97(1),233-255.

Ogbu, J. (1988). Minority education and caste. New York: Academic Press.

Olsen L & Jaramillo, A. (1999.) Turning the tide of exclusion: A guide for educators and advocates of immigrant students. California Tomorrow.

Rivera, C.; Stansfield, C.W; Scialdone, L; Sharkey, M. (2000). An analysis of state policies for the inclusion and accommodation of English Language Learners in state assessment programs during 1998-1999. Center for Equity and Excellence in Education: George Washington University.

Thomas, W. & Collier, V. (1996). *Language minority student achievement and program effectiveness.* NABE News, 33-35.